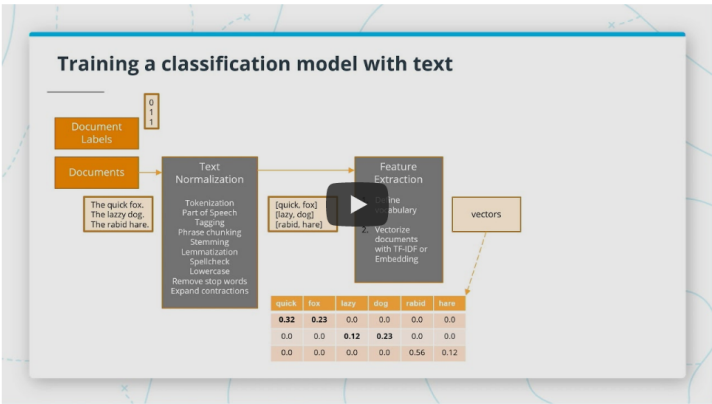


Text Classification

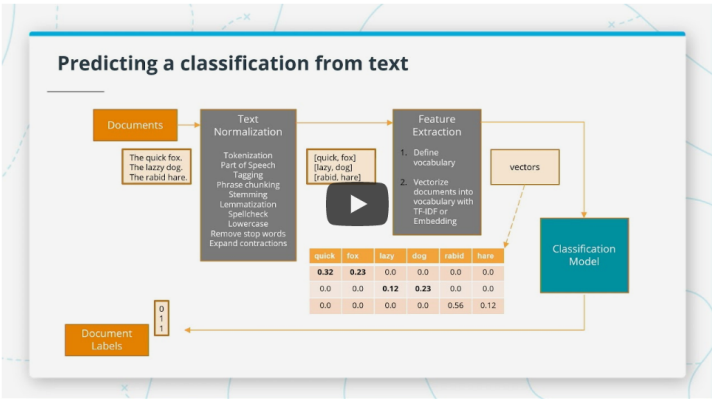


Remember that before we can do text classification, the text first needs to be translated into some kind of numerical representation—a process known as **text embedding**. The resulting numerical representation, which is usually in the form of vectors, can then be used as an input to a wide range of classification algorithms.

Training a Classification Model with Text



As we mentioned in the video, an important part of the pipeline is the process of vectorizing the text, using a technique such as [Term Frequency-Inverse Document Frequency \(TF-IDF\) vectorization](#). We discussed text vectorization in some detail back in the introduction lesson when we described text data —so you may want to go back to that section if you feel like you need a review of the concept.



QUIZ QUESTION

Which of the following is not a typical step in model training for text classification?

- ☒ Document labeling
- ☐ Text normalization
- ☐ Supervised learning model training
- ☐ Feature extraction

SUBMIT

NEXT